INTERNATIONAL STANDARD

ISO/IEC 10646

First edition
2003-12-15

AMENDMENT 3
2008-02-15

Information technology — Universal Multiple-Octet Coded Character Set (UCS) —

AMENDMENT 3: Lepcha, Ol Chiki, Saurashtra, Vai and other characters

Technologies de l'information — Jeu universel de caractères codés sur plusieurs octets (JUC) —

AMENDEMENT 3: Lepcha, Ol Chiki, Saurashtra, Vai et autres caractères

Reference number
ISO/IEC10646:2003/Amd.3:2008(E)

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of the joint technical committee is to prepare International Standards. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national bodies casting a vote.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

Amendment 3 to ISO/IEC 10646:2003 was prepared by Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 2, *Coded character sets*.

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of the joint technical committee is to prepare International Standards. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national bodies casting a vote.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

Amendment 3 to ISO/IEC 10646:2003 was prepared by Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 2, *Coded character sets*.

Information technology — Universal Multiple-Octet Coded Character Set (UCS) —

AMENDMENT 3: Lepcha, Ol Chiki, Saurashtra, Vai, and other characters

Page 2, Clause 3 Normative references

Add the following normative reference

Unicode Technical Standard, UTS#37, Ideographic Variation Database, Version 1.0, January 2006.

Page 2, Clause 4 Terms and definitions

Replace the first paragraph of the Collection definition (4.12) with the following:

A numbered and named set of entities. For a non extended collection, these entities consist only of those coded characters whose code positions lie within one or more identified ranges (see also 4.21 for extended collection).

Insert the following before the current 4.16 and update accordingly all following term numbers and cross references accordingly.

4.16 Control character

A control function the coded representation of which consists of a single code position.

NOTE – Although control characters are often 'named' using terms such as DELETE, FORM FEED, ESC, these qualifiers do not correspond to formal character names. See clause 15 for a list of the long names used by ISO/IEC 6429 in association with the control characters.

*In the definition of Control function, replace "*has a coded representation consisting of one or more octets*" with "*is represented by a CC-data-element*".*

Insert the following before the current 4.20 and update accordingly all following term numbers and cross references accordingly.

4.21 Extended Collection

A collection for which the entities can also consist of sequences of code positions that are in normalization form NFC (see clause 25). The sequences of code positions are referenced by Named UCS Sequence Identifiers (NUSI) listed in clause 29 (see also 4.12).

NOTE – Some collections such as 3 LATIN EXTENDED-A, 4 LATIN EXTENDED-B, 15 ARABIC EXTENDED, and many more, have the term 'extended' in their name. This does not make them extended collections.

Page 4 Clause 5 General structure of the UCS

Delete the two last paragraphs about transformation formats.

Page 8, Sub-clause 6.3 Octet order

Replace the paragraph as follows:

In the canonical form of the coded character set, the sequence of the octets that represent a character, and the most significant and least significant ends of it, shall be maintained as shown above.

Other forms of coded representation such as UTF-16 and UTF-8, have their own sequence of octets as indicated in annex C and D respectively.

The order of octets in the coded representation form may be determined by the usage of a signature at the start of the data stream (see annex H), by the declaration of features identification (see 16.1), or by the usage of specific transformation formats such as UTF-16BE, UTF-16LE (see annex C), UTF-32BE, and UTF-32LE (see 13.2).

Page 9, Sub-clause 6.6 UCS Sequence Identifiers

In the note replace "subset and collection" *with* "subset".

Page 9, Clause 8 The Basic Multilingual Plane

Replace the second paragraph with the following:

Code positions 0000 to 001F, 007F to 009F in the BMP are reserved for control characters (see clause 15).

Page 11, Clause 13 Coded representation forms of the UCS

Replace the first paragraph with the following text:

ISO/IEC 10646 provides eight alternative forms of coded representation of characters. Four of these forms are specified in this clause (UCS-2, UCS-4, UTF-32BE, and UTF-32LE). Three others are specified in annex C (UTF16, UTF-16BE, and UTF-16LE). Finally, UTF-8 is specified in annex D.

Page 11, Sub-clause 13.1 Two-octet BMP form (UCS-2)

Add the following sentence at the end of the second paragraph:

For serialization purpose, a signature may be used (see annex H).

Page 11, Sub-clause 13.2 Four-octet canonical form (UCS-4)

*Replace sub-clause header with '***Four-octet canonical forms (UCS-4, UTF-32BE, and UTF-32LE)***'.*

In first paragraph, replace 'The canonical form permits' with 'These canonical forms permit'.

In the second paragraph, add 'UCS-4' after 'four-octet canonical form'.

Replace NOTE 2, 3, and 4 with following:

UCS-4 is also referred to as UCS Transformation Format (UTF-32). For serialization purpose, a signature may be used (see annex H)

> NOTE 2 – UTF-32 was originally specified by the Unicode Standard and restricted to the code positions in Planes 00 to 10 (U+0000 to U+10FFF). Because code positions in all other planes are now permanently reserved, UCS-4 and UTF-32 can be used interchangeably for all assigned characters.

Two additional four-octet UCS Transformation Formats are specified for serialization purpose.

1. UTF-32BE: in the ordering of octets the more significant octets precede the less significant octets, as specified in clause 6.2, and no signatures appear;
2. UTF-32LE: in the ordering of octets the less significant octets precede the more significant octets, and no signatures appear.

Page 11, Clause 15 Use of control function with the UCS

Renumber the current "NOTE 1" and add the following note at the end of the clause:

NOTE 2 – The following list provides the long names from ISO/IEC 6429 used in association with the control characters.
0000 NULL
0001 START OF HEADING
0002 START OF TEXT
0003 END OF TEXT
0004 END OF TRANSMISSION
0005 ENQUIRY
0006 ACKNOWLEDGE
0007 BELL
0008 BACKSPACE
0009 CHARACTER TABULATION
000A LINE FEED
000B LINE TABULATION
000C FORM FEED
000D CARRIAGE RETURN
000E SHIFT-OUT
000F SHIFT-IN
0010 DATA LINK ESCAPE
0011 DEVICE CONTROL ONE
0012 DEVICE CONTROL TWO
0013 DEVICE CONTROL THREE
0014 DEVICE CONTROL FOUR
0015 NEGATIVE ACKNOWLEDGE
0016 SYNCHRONOUS IDLE
0017 END OF TRANSMISSION BLOCK
0018 CANCEL
0019 END OF MEDIUM
001A SUBSTITUTE
001B ESCAPE
001C INFORMATION SEPARATOR FOUR
001D INFORMATION SEPARATOR THREE
001E INFORMATION SEPARATOR TWO
001F INFORMATION SEPARATOR ONE
007F DELETE
0082 BREAK PERMITTED HERE
0083 NO BREAK HERE
0084 INDEX
0085 NEXT LINE
0086 START OF SELECTED AREA
0087 END OF SELECTED AREA
0088 CHARACTER TABULATION SET
0089 CHARACTER TABULATION WITH JUSTIFICATION
008A LINE TABULATION SET
008B PARTIAL LINE FORWARD
008C PARTIAL LINE BACKWARD
008D REVERSE LINE FEED
008E SINGLE-SHIFT TWO
008F SINGLE-SHIFT THREE
0090 DEVICE CONTROL STRING
0091 PRIVATE USE ONE
0092 PRIVATE USE TWO
0093 SET TRANSMIT STATE
0094 CANCEL CHARACTER
0095 MESSAGE WAITING
0096 START OF GUARDED AREA
0097 END OF GUARDED AREA
0098 START OF STRING
009A SINGLE CHARACTER INTRODUCER
009B CONTROL SEQUENCE INTRODUCER
009C STRING TERMINATOR
009D OPERATING SYSTEM COMMAND
009E PRIVACY MESSAGE
009F APPLICATION PROGRAM COMMAND

The control character 0084 INDEX has been removed from ISO/IEC 6492:1992. In addition, the control characters 000E and 000F are named SHIFT-OUT and SHIFT-IN respectively

in 7-bit environment and LOCKING-SHIFT ONE and LOCKING-SHIFT ZERO respectively in 8-bit environment.

Page 15, Sub-Clause 20.4 Variation selectors

Replace the sub-clause title and the first four paragraphs of the sub-clause, ending with 'composite characters will be defined.' with the following text:

20.4 Variation selectors and variation sequences

Variation selectors are a specific class of combining characters immediately following a non decomposable base character and which indicate a specific variant form of graphic symbol for that character. A decomposable character is a character for which there exists an equivalent composite sequence. The character sequence consisting of a non decomposable base character followed by a variation selector is called a variation sequence.

NOTE 1 – Some variation selectors are specific to a script, such as the Mongolian free variation selectors, others are used with various other base characters such as the mathematical symbols.

Only the variation sequences defined or referenced in this clause indicate a specific variant form of graphic symbol; all other such sequences are undefined. Furthermore, variation selectors following other base characters and any non-base characters have no effect on the selection of the graphic symbol for that character.

No variation sequences using characters from VARIATION SELECTOR-2 to VARIATION SELECTOR-16 are defined at this time. Variations sequences composed of a unified ideograph as the base character and one of VARIATION SELECTOR-17 to VARIATION SELECTOR-256 from the Supplementary Special-purpose Plane (SSP) are registered in the Ideographic Variation Database defined by Unicode Technical Standard #37.

NOTE 2 – The Ideographic Variation Database is currently empty. When entries are registered, these variation sequences will be referenced by this standard.

In note 5, replace the referenced link to the Unicode character database with the following:

http://www.unicode.org/Public/5.0.0/ucd/StandardizedVariants.html

Page 18, Clause 25 Normalization forms

Add the following text at the end of the first paragraph:

There are four normalization forms:

1. Normalization Form D (NFD) which is a canonical decomposition,
2. Normalization Form C (NFC) which is a canonical decomposition followed by canonical composition,
3. Normalization Form KD (NFKD) which is a compatibility decomposition,
4. Normalization Form KC (NFKC) which is a compatibility decomposition followed by canonical composition.

Replace the first note as follows:

NOTE 1 – The result of applying any of these normalization forms onto a CC-data-element is intended to stay stable over time. It means that the normalized representation of a CC-data-element consisting of characters assigned in this version of the standard remains normalized even when the standard is amended.

Page 20, Sub-clause 26.1 Hangul syllable composition method

Remove the word 'character' in the three parenthetical notations in the first paragraph.

Page 23, Clause 28 Character names and annotations

*Remove the header 28.1 General and move up by one level the headers corresponding from sub-clauses 28.1.1 to 28.1.5. This makes them sub-clauses 28.1, 28.2, 28.3, 28.4, and 28.5. The existing sub-clauses 28.2 and 28.3 becomes 28.6 and 28.7. Within the previous sub-clause '*28.1.4 Name Uniqueness*', now sub-clause '*28.4 Name Uniqueness*', the titles '*Block names*', '*Collection names*', '*Character names and named UCS sequence identifiers*', and '*Determining uniqueness*' become sub-clauses 28.4.1, 28.4.2, 28.4.3, and 28.4.4 respectively.*

Page 25, Clause 29 Named UCS Sequence Identifiers

*In first paragraph, replace '*A named UCS Sequence Identifier (USI)*' with '*A Named UCS Sequence Identifier (NUSI)*'.*

Add the following paragraph after the note:

The USI value corresponding to each NUSI is written using the coded representation determined by the normalization form NFC (see clause 25). Each named UCS sequence has a unique code representation.

Move the paragraph starting with 'All the allowed…' and ending with '…are undefined' immediately after that new text.

In the list, delete the '<00E1, 0328>' *entry.*

Insert the following entries in the list:

<0104, 0301> LATIN CAPITAL LETTER A WITH OGONEK AND ACUTE

<0105, 0301> LATIN SMALL LETTER A WITH OGONEK AND ACUTE

<0104, 0303> LATIN CAPITAL LETTER A WITH OGONEK AND TILDE
<0105, 0303> LATIN SMALL LETTER A WITH OGONEK AND TILDE
<0118, 0301> LATIN CAPITAL LETTER E WITH OGONEK AND ACUTE
<0119, 0301> LATIN SMALL LETTER E WITH OGONEK AND ACUTE
<0118, 0303> LATIN CAPITAL LETTER E WITH OGONEK AND TILDE
<0119, 0303> LATIN SMALL LETTER E WITH OGONEK AND TILDE
<0116, 0301> LATIN CAPITAL LETTER E WITH DOT ABOVE AND ACUTE
<0117, 0301> LATIN SMALL LETTER E WITH DOT ABOVE AND ACUTE
<0116, 0303> LATIN CAPITAL LETTER E WITH DOT ABOVE AND TILDE
<0117, 0303> LATIN SMALL LETTER E WITH DOT ABOVE AND TILDE
<0069, 0307, 0300>LATIN SMALL LETTER I WITH DOT ABOVE AND GRAVE
<0069, 0307, 0303>LATIN SMALL LETTER I WITH DOT ABOVE AND TILDE
<012E, 0301> LATIN CAPITAL LETTER I WITH OGONEK AND ACUTE
<012F, 0307, 0301> LATIN SMALL LETTER I WITH OGONEK AND DOT ABOVE AND ACUTE
<012E, 0303> LATIN CAPITAL LETTER I WITH OGONEK AND TILDE
<012F, 0307, 0303> LATIN SMALL LETTER I WITH OGONEK AND DOT ABOVE AND TILDE
<004A, 0303> LATIN CAPITAL LETTER J WITH TILDE
<006A, 0307, 0303> LATIN SMALL LETTER J WITH DOT ABOVE AND TILDE
<004C, 0303> LATIN CAPITAL LETTER L WITH TILDE
<006C, 0303> LATIN SMALL LETTER L WITH TILDE
<004D, 0303> LATIN CAPITAL LETTER M WITH TILDE
<006D, 0303> LATIN SMALL LETTER M WITH TILDE
<0052, 0303> LATIN CAPITAL LETTER R WITH TILDE
<0072, 0303> LATIN SMALL LETTER R WITH TILDE
<0172, 0301> LATIN CAPITAL LETTER U WITH OGONEK AND ACUTE
<0173, 0301> LATIN SMALL LETTER U WITH OGONEK AND ACUTE
<0172, 0303> LATIN CAPITAL LETTER U WITH OGONEK AND TILDE
<0173, 0303> LATIN SMALL LETTER U WITH OGONEK AND TILDE
<016A, 0301> LATIN CAPITAL LETTER U WITH MACRON AND ACUTE
<016B, 0301> LATIN SMALL LETTER U WITH MACRON AND ACUTE
<016A, 0303> LATIN CAPITAL LETTER U WITH MACRON AND TILDE
<016B, 0303> LATIN SMALL LETTER U WITH MACRON AND TILDE

Page 30-1348 Clause 33, Code Tables and list of character names

1. Modifications to existing tables

Insert the additional character glyphs and names at the indicated positions in the tables given below, the character names replacing the existing entries which read "(This position shall not be used)". The table numbers are affected by the insertion of new tables (see below) preceding these modified tables. (The table numbers corresponding to the first edition of ISO/IEC 10646:2003 are mentioned in parenthesis.)

Plane 00

Table 2 - Row 00: Basic Latin (2)
Table 3 - Row 00: Latin-1 Supplement (3)
Table 7 - Row 02: IPA Extensions (7)
Table 9 - Row 03: Combining Diacritical Marks (3)
Table 10 - Row 03: Greek and Coptic (10)
Table 11 - Row 04: Cyrillic (11)
Table 12 - Row 04: Cyrillic (12)
Table 16 - Row 06: Arabic (16)
Table 19 - Row 07: Arabic Supplement
Table 22 - Row 09: Devanagari (20)
Table 24 - Row 0A: Gurmukhi (22)
Table 27 - Row 0B: Tamil (25)
Table 28 - Row 0C: Telugu (26)
Table 30 - Row 0D: Malayalam (28)
Table 34 - Row 0F: Tibetan (32)
Table 35 - Row 0F: Tibetan (33)
Table 36 - Row 10: Myanmar (34)
Table 56 - Row 18: Mongolian (53)
Table 68 - Row 1D: Combining Diacritical Marks Supplement
Table 70 - Row 1E: Latin Extended Additional (59)
Table 81 - Row 22: Mathematical Operators (70)
Table 90 - Row 26: Miscellaneous Symbols (79)
Table 91 - Row 26: Miscellaneous Symbols (80)
Table 94 - Row 27: Miscellaneous Mathematical Symbols-A (83)
Table 102 - Row 2B: Miscellaneous Symbols and Arrows (91)
Table 104 - Row 2C: Latin Extended-C
Table 109 - Row 2E: Supplemental Punctuation
Table 121 - Row 31: CJK Strokes
Table 138 - Row A7: Modifier Tone Letters
Table 139 - Row A7: Latin Extended-D

Add a '' (denoting an entry in annex P) after 000E and 000F in Table 02, and 1DA6, 1DAB, 1DB0, and 1DB8 in Table 67. The same notation is added in annex G.*

Plane 01

Table 221 - Row D1: Musical Symbols (177)

Add a '' (denoting an entry in annex P) after 1D13A in Table 221. The same notation is added in annex G.*

These tables contain new characters and names at the following code positions:

0370-0373, 0376-0377, 03CF, 0606-060A, 076E-077D, 0971-0972, 097B-097C, 097E-097F, 0A51, 0A75, 0BD0, 0C3D, 0C58-0C59, 0C62-0C63, 0C78-0C7F, 0D3D, 0D44, 0D62, 0D70-0D75, 0D79, 0F6B-0F6C, 0FCE, 0FD2-0FD4, 1028, 102B, 1033-1035, 103A-103F, 105A-1064, 18AA, 1DCB-1DE6, 1E9C-1E9D, 1E9F, 1EFA-1EFF, 269D, 26B3-26BC, 27EC-27ED, 2B30-2B44, 2B47-2B4C, 2C6D-2C6F, 2C71-2C73, 2C78-2C7D, 2E18, 31D0-31E3, A71B-A71F, A722-A788, 1D129

new graphic symbols and annotations at the following code positions:

0000-001F, 007F-009F

and updated graphic symbols at the following code positions:

027F, 0285, 0340-0341, 03E0-03EF, 0460, 0478-0479, 047C-047E, 0485-0486, 1039, 104E, 22C4, 2626

2. New tables

Insert the following additional tables and adjust the numbering of the existing tables that follow. When correctly applied, all tables will be arranged by ascending code position.

Plane 00

Table 63 - Row 1B: Sundanese
Table 64 - Row 1C: Lepcha
Table 65 - Row 1C: Ol Chiki
Table 136 - Row A5: Vai
Table 137 - Row A5-A6: Vai
Table 142 - Row A8: Saurashtra
Table 143 - Row A9: Kayah Li
Table 144 - Row A9: Rejang

Plane 01

Table 198 - Row 01: Phaistos Disc
Table 199 - Row 02: Lycian
Table 200 - Row 02: Carian
Table 210 - Row 09: Lydian

These tables add new characters and names at the following code positions:

1B80-1BAA, 1BAE-1BB9, 1C00-1C37, 1C3B-1C49, 1C4D-1C4F, 1C50-1C7F, A500-A62B, A880-A8C4, A8CE-A8D9, A900-A92F, A930-A953, A95F, 101D0-101FD, 10280-1029C, 102A0-102D0, 10920-10939, 1093F

Page 1349, Annex A.1 Collections of coded graphic characters

*Replace the clause title with "***A.1 Collections of graphic characters***".*

Add a '' (for fixed collections) to the following collection:*

125 MODIFIER TONE LETTERS

In the list of collection numbers and names, after

132 PHAGS-PA

insert new entries as follows:

133 SUNDANESE	1B80-1BBF
134 LEPCHA	1C00-1C4F
135 OL CHIKI	1C50-1C7F *
136 VAI	A500-A63F
137 SAURASHTRA	A880-A8DF
138 KAYAH LI	A900-A92F *
139 REJANG	A930-A95F

after

1022 COUNTING ROD NUMERALS

insert new entries as follows:

1023 PHAISTOS DISC	101D0-101FF
1024 LYCIAN	10280-1029F
1025 CARIAN	102A0-102DF
1026 LYDIAN	10920-1093F

Page 1351, Annex A.1

In the list of collections numbers and names, after

207 IDEOGRAPHIC DESCRIPTION CHARACTERS

insert the new entry:

208	CONTROL CHARACTERS	0000-001F 0007F-009F

Page 1351, Annex A.1

In the list of collections numbers and names, replace the text after

3002 ALTERNATE FORMAT CHARACTERS

and before the second note with the following:

The following specify collections that represented the whole UCS when they were created.

299	(This collection number shall not be used, see A.3.2.)	
301	BMP-AMD.7	see A.3.1 *

302	BMP SECOND EDITION	see A.3.3 *
303	UNICODE 3.1	see A.6.1 *
304	UNICODE 3.2	see A.6.2 *
305	UNICODE 4.0	see A.6.3 *
306	UNICODE 4.1	see A.6.4 *
307	UNICODE 5.0	see A.6.5 *
340	COMBINED FIRST EDITION	see A.3.4 *
10646	UNICODE	0000-FDCF FDF0-FFFD 10000-1FFFD 20000-2FFFD 30000-3FFFD 40000-4FFFD 50000-5FFFD 60000-6FFFD 70000-7FFFD 80000-8FFFD 90000-9FFFD A0000-AFFFD B0000-BFFFD C0000-CFFFD D0000-DFFFD E0000-EFFFD F0000-FFFFD 100000-10FFFD

NOTE 1 – The UNICODE collection incorporates all characters currently encoded in the standard.

The following collections only contain CJK ideographs.

370	IICORE	see A.4.1 *
371	JIS2004 IDEOGRAPHICS EXTENSION *	see A.4.2
372	JAPANESE IDEOGRAPHICS SUPPLEMENT *	see A.4.3
380	CJK UNIFIED IDEOGRAPHS-2001	3400-4DB5 * 4E00-9FA5 FA0E-FA0F FA11 FA13-FA14 FA1F FA21 FA23-FA24 FA27-FA29 20000-2A6D6
381	CJK COMPATIBILITY IDEOGRAPHS-2001 *	F900-FA0D FA10 FA12 FA15-FA1E FA20 FA22 FA25-FA26 FA2A-FA6A 2F800-2FA1D
382	CJK UNIFIED IDEOGRAPHS-2005	Collection 380* 9FA6-9FBB
383	CJK COMPATIBILITY IDEOGRAPHS-2005 *	Collection 381 FA70-FAD9

The following specify other collections, including extended collections.

270	COMBINING CHARACTERS characters specified in annex B.1	
271	COMBINING CHARACTERS B-2 characters specified in annex B.2	
281	MES-1	see A.5.1 *
282	MES-2	see A.5.2 *
283	MODERN EUROPEAN SCRIPTS	see A.5.3 *
284	CONTEMPORARY LITHUANIAN LETTERS *	see A.5.4
285	BASIC JAPANESE	see A.5.5 *
286	JAPANESE NON IDEOGRAPHICS EXTENSION *	see A.5.6
287	COMMON JAPANESE	see A.5.7 *
300	BMP	0000-D7FF E000-FFFD
400	(This collection number shall not be used, see Note 3.)	
401	PRIVATE USE PLANES-0F-10	G=00, P=0F-10
500	(This collection number shall not be used, see Note 3.)	
1000	SMP	10000-1FFFD
1900	SMP COMBINING CHARACTERS SMP characters specified in annex B.1	
2000	SIP	20000-2FFFD
3000	SSP	E0000-EFFFD

The following specify collections which are the union of particular collections defined above.

63	ALPHABETIC PRESENTATION FORMS	Collections 104-105
250	GENERAL FORMAT CHARACTERS	Collections 200-203
251	SCRIPT-SPECIFIC FORMAT CHARACTERS	Collections 204-206
4000	UCS PART-2	Collections 1000, 2000, 3000

Page 1351, Annex A.1

Under Note 2, after "128", add ",133, 134, 137, 138, 139".

Under Note 2, after "1017", replace "and 1018" with "1018, and 1023".

In the alphabetical list of keywords in Note 4, add collection "307" to the entry "Unicode".

In the alphabetical list of keywords in Note 4, insert the following entries:

Carian	1025
Kayah Li	138
Lepcha	134
Lycian	1024
Lydian	1026
Ol Chiki	135
Phaistos Disc	1023
Rejang	139
Saurashtra	137
Sundanese	133
Vai	136

Page 1352, Annex A.2.1

In the list of blocks in the BMP, insert the following new entries:

SUNDANESE	1B80-1BBF
LEPCHA	1C00-1C4F
OL CHIKI	1C50-1C7F
VAI	A500-A63F
SAURASHTRA	A880-A8DF
KAYAH LI	A900-A92F
REJANG	A930-A95F

Page 1353, Annex A.2.2

In the list of blocks in the SMP, insert the following new entries:

PHAISTOS DISC	101D0-101FF
LYCIAN	10280-1029F
CARIAN	102A0-102DF
LYDIAN	10920-1093F

Page 1353, Annex A.3

Replace header title and add paragraph as follows:

A.3 Fixed collections of the whole UCS (except Unicode collections)

The following collections contain the whole UCS assigned character content as it was when they were created. The Unicode collections are described in A.6.

Move annex A.5.1 (340 COMBINED FIRST EDITION) to A.3.4.

Page 1355, Annex A.4

Replace header title as follows:

A.4 CJK collections

Remove first paragraph and note.

Move all existing A.4 annexes to A.5 (see below).

Move annex A.5.2 (370 IICORE) to A.4.1.

Add the following annexes:

A.4.2 371 JIS2004 IDEOGRAPHICS EXTENSION

371 The fixed collection JIS2004 IDEOGRAPHICS EXTENSION consists of all level 3 and level 4 CJK characters defined in JIS X 0213:2004.

NOTE 1 – Given its large size (3695 characters) and the large number of sparse ranges, the collection is not specified by Rows/Positions but instead by a linked content.

The content linked to is a plain text file, using ISO/IEC 646-IRV characters with LINE FEED as end of line mark, that specifies, after a 3-lines header, as many lines as characters in the collection; each containing the following information in fixed length field:

- BMP or SIP code position (`0hhhh`), (`2hhhh`), normative.

The format definition uses '`h`' as a hexadecimal unit. Digits between parentheses appear as shown.

Click on this highlighted text to access the reference file.

NOTE 2 – The content is also available as a separate viewable file in the same file directory as this document. The file is named: "JIEx.txt".

A.4.3 372 JAPANESE IDEOGRAPHICS SUPPLEMENT

372 The fixed collection JAPANESE IDEOGRAPHICS SUPPLEMENT consists of all CJK characters defined in JIS X 0212:1990. It contains 5801 characters.

NOTE – 2742 characters are common between the collections 371 and 372.

The code positions of this collection are identified by the J1 Kanji J sources in the Source Reference file for CJK Unified Ideographs (CJKU_SR.txt). See clause 27.1 for further details.

Page 1356, Annex A.5

Replace header title and add paragraph as follows:

A.5 Other collections

The collections specified within this clause address the referencing need of users community. Characters may be from different writing systems and may be coded in different planes. It includes collection for users community from Lithuania, Japan and Europe as a whole.

NOTE – The acronym MES used in collection names below indicates Multilingual European Subset.

Insert former annex A.4.1 (281 MES-1) as annex A.5.1. Insert 'Plane 00' before the 'Rows/Positions' line.

Insert former annex A.4.2 (282 MES-2) as annex A.5.2. Insert 'Plane 00' before the 'Rows/Positions' line.